

ℓ_0 -Optimization for Channel and DOA Sparse Estimation

Adilson Chinatto*, Emmanuel Soubies[‡], Cynthia Junqueira*, João M. T. Romano*,
Pascal Larzabal[†], Jean-Pierre Barbot[†] and Laure Blanc-Féraud[‡]

*School of Electrical and Computer Engineering, University of Campinas, Campinas, Brazil

[†]Laboratoire SATIE, École Normale Supérieure de Cachan, Cachan, France

[‡]Laboratoire I3S, Université de Nice Sophia Antipolis, Nice, France

Abstract—This paper is devoted to two classical sparse problems in array processing: Channel estimation and DOA estimation. It is shown after some background and some recent results in ℓ_0 optimization how this latter can be used, at the same computational cost, in order to obtain improvement in comparison with ℓ_1 optimization for sparse estimation.

I. INTRODUCTION

Sparse optimization has been a topic of great interest in recent years encountering many applications in fields as coding, estimation, source separation, compressive sampling, image decomposition among others [1]. In all these applications, the problem is to search a sparse solution of an underdetermined system of linear equations. One example is the problem of finding a representation of the unknown signal as a linear combination of a few redundant dictionary atoms.

In general, sparse optimization is an ill-posed inverse problem which can be regularized by the introduction of sparsity constraints based on the ℓ_0 pseudo-norm. However, such a regularization leads to nonconvex combinatorial problems known to be NP-hard [2]. A popular alternative is then to relax the ℓ_0 pseudo-norm by the convex ℓ_1 -norm to take benefit from the well known convex optimization framework whose solvers has been developed since 1960's. The resulting optimization problem is also known as *Basis Pursuit* (BP) or *Basis Pursuit De-Noising* (BPDN) [3] and is usually addressed thanks to linear programming algorithms. Moreover, iterative algorithms, as the *Iterative Shrinkage-Thresholding Algorithm* (ISTA) together with its variants [4], have shown their efficiency for such ℓ_1 regularized problems. Under some conditions on the mixture matrix, such as the Restricted Isometry Property (RIP) [8], exact recovery of sufficiently sparse signals can be guaranteed by ℓ_1 minimization. However these conditions are generally too restrictive for many applications. Greedy algorithms have also been widely investigated in the context of ℓ_0 -optimization. Among them, we can cite the *Matching Pursuit* (MP) algorithm, initially introduced in [5], which has been later improved with *Orthogonal Matching Pursuit* (OMP) [6] and more recently with *Single Best Replacement* (SBR) [7]. Finally, numerous nonconvex *continuous* penalties have been proposed to approach the ℓ_0 pseudo-norm. The main motivation was to better promote sparsity than the ℓ_1 -norm while getting the continuity unlike the ℓ_0 penalty. Moreover, efficient algorithms have been dedicated to deal with such

penalties. The following work takes place within this context of nonconvex *continuous* penalties and iterative algorithms to solve underlying minimization problems.

Contributions and Outline: Based on the Forward-Backward Splitting (FBS) algorithm [11], [13] we discuss and compare the well known hard thresholding rule, leading to the Iterative Hard Thresholding algorithm (IHT) [9] and minimizing directly the ℓ_0 penalized functional, with a *continuous* threshold associated to a recently introduced Continuous Exact ℓ_0 Penalty (CEL0) [10]. This penalty preserves global minimizers of the initial ℓ_0 penalized functional. Comparisons with the classical soft thresholding, associated to ℓ_1 -norm minimization, are also provided. These algorithms are applied to sparse channel and direction of arrival (DOA) estimation problems. In the case of channel estimation, the objective is to enhance the estimation resolution by using the sparsity characteristic of telecommunications channel. In the DOA estimation case, the challenge is to estimate correctly the incident angles of signals impinging an antenna array without using conventional techniques as those based on subspace methods, that demand high computational loads. The paper is organized as follows: Section II presents the FBS algorithm for convex optimization and its application to ℓ_1 minimization; Section III brings recent results in sparse optimization using ℓ_0 -norm minimization; in Section IV the concepts of nonconvex iterative optimization are applied and compared on the sparse channel estimation problem; Section V presents results of nonconvex optimization for the DOA estimation problem; finally, Section VI brings work's conclusions and future perspectives.

Notations: lowercase italic letters means scalars; lowercase bold letters are vectors and uppercase bold letters are matrices; the transpose and hermitian operators are represented by $(\cdot)^T$ and $(\cdot)^H$, respectively; euclidean (ℓ_2) norm will be denoted $\|\cdot\|$ while other ℓ_p -norms ($p \neq 2$) will be specified with a subscript as $\|\cdot\|_p$ for $p > 0$; special case when $p = 0$ is called ℓ_0 pseudo-norm and $\|\mathbf{x}\|_0$ counts all non-null entries of \mathbf{x} . Finally $\mathbb{1}_E(x) = \{1 \text{ if } x \in E; 0 \text{ otherwise}\}$ and $x_+ = \{x \text{ if } x > 0; 0 \text{ otherwise}\}$.

II. FORWARD-BACKWARD SPLITTING (FBS) ALGORITHM

Let us consider the following optimization problem,

$$\underset{\mathbf{x} \in \mathbb{C}^{N \times 1}}{\text{minimize}} f(\mathbf{x}) + g(\mathbf{x}), \quad (1)$$

where $f(\mathbf{x})$ and $g(\mathbf{x})$ are proper lower-semicontinuous convex functions and $f(\mathbf{x})$ is differentiable with a Lipschitz continuous gradient. Aiming to solve this kind of problem, a simple algorithm, called *Forward-Backward Splitting* (FBS) has been introduced in the context of signal processing [11]. Following [11, Proposition 3.1], solutions of (1) are characterized by the fixed point equation,

$$\mathbf{x} = \text{prox}_{\gamma g}(\mathbf{x} - \gamma \nabla f(\mathbf{x})), \quad (2)$$

for $\gamma > 0$ and where the *proximity operator* is defined by,

$$\text{prox}_g(\mathbf{u}) := \arg \min_{\mathbf{x} \in \mathbb{C}^{N \times 1}} \left(g(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{u}\|^2 \right), \quad (3)$$

which admits a unique solution for every $\mathbf{u} \in \mathbb{C}^{N \times 1}$ [12]. Then, equation (2) enables the possibility of pursuit the fixed-point iteratively by performing iterations as,

$$\mathbf{x}^{(k+1)} = \text{prox}_{\gamma g}(\mathbf{x}^{(k)} - \gamma \nabla f(\mathbf{x}^{(k)})), \quad (4)$$

where k is the iteration number and $\gamma \in (0, \frac{2}{L})$ with L the Lipschitz constant of the gradient of f [11]. Finally, the *Forward-Backward Splitting* nomenclature comes from an explicit forward gradient step using the function f and a implicit backward step using the function g .

When $g(\cdot) = \lambda \|\cdot\|_1$, the proximity operator in (4) is defined by the following soft thresholding rule,

$$\text{prox}_{\gamma \lambda \|\cdot\|_1}(\mathbf{u}) = \left[\frac{u_i}{|u_i|} \max(0, |u_i| - \gamma \lambda) \right]_{i \in \{1, \dots, N\}}, \quad (5)$$

and (4) leads to the well-known ISTA [4].

III. RECENT RESULTS IN ℓ_0 -OPTIMIZATION

Let $\mathbf{x} \in \mathbb{C}^{N \times 1}$ be a sparse signal vector and $\mathbf{A} \in \mathbb{C}^{M \times N}$, $M \ll N$, a mixture matrix containing the dictionary that produces a set of non-sparse observations $\mathbf{y} \in \mathbb{C}^{M \times 1}$ corrupted by a noise $\mathbf{n} \in \mathbb{C}^{M \times 1}$. This set of linear equations can then be expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (6)$$

where the noise \mathbf{n} is considered to be gaussian with zero mean and σ^2 variance. In order to recover the sparse signal \mathbf{x} , one can search for a solution of the following least-squares penalized minimization problem,

$$\hat{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \mathbb{C}^{N \times 1}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{x}\|_0, \quad (7)$$

whose estimation performance depends on the penalization parameter λ , a compromise between data-fidelity and sparsity. This problem belongs to the so called class of ℓ_0 -optimization problems known to be NP-hard [2].

Although the substitution of ℓ_0 - by ℓ_1 -norm in (7) enables the use of convex optimization techniques such as the FBS algorithm presented in Section II, *exact recovery* of the sparse vector \mathbf{x} using this relaxation is only guaranteed under some restrictions on \mathbf{A} such as RIP criteria [8]. Consequently, good recovery will be performed by ℓ_1 minimization only if \mathbf{x} is sufficiently sparse and/or if the columns of \mathbf{A} are

sufficiently uncorrelated. In practical applications, however, these conditions are too restrictive.

Recently, some techniques have been developed to perform a direct optimization of problem (7) through iterative algorithms with local convergence guaranties. Problem (7) can thus be solved by the IHT algorithm [9], a special case of FBS for $g(\cdot) = \lambda \|\cdot\|_0$ where the proximity operator in (4) is the hard threshold given by,

$$\text{prox}_{\gamma \lambda \|\cdot\|_0}(\mathbf{u}) = \left[u_i \mathbb{1}_{\{|u_i| > \sqrt{2\gamma\lambda}\}} \right]_{i \in \{1, \dots, N\}}. \quad (8)$$

FBS is a very simple algorithm iterating between a gradient step and a thresholding. Note that convergence of FBS in the nonconvex case has been recently established in [13]. For $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$ (i.e. problem (7)), we have

$$\nabla f(\mathbf{x}) = \mathbf{A}^H (\mathbf{A}\mathbf{x} - \mathbf{y}). \quad (9)$$

Although IHT presents an interesting iterative solution for ℓ_0 -optimization, a major drawback of this algorithm comes from its sensitivity to the initialization. Indeed, from a “bad” initial guess, it can converges to a non-sparse local minimizer of the ℓ_2 - ℓ_0 criteria (7). In order to be more robust against the initialization while preserving the convergence to a local minimizer of ℓ_2 - ℓ_0 , an alternative consists in replacing the ℓ_0 -norm by the CEL0 penalty, recently introduced in [10] and leading to a reformulation of problem (7) as,

$$\hat{\mathbf{x}} \in \arg \min_{\mathbf{x} \in \mathbb{C}^{N \times 1}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \Phi_{\text{CEL0}}(\mathbf{x}), \quad (10)$$

where,

$$\Phi_{\text{CEL0}}(\mathbf{x}) := N\lambda - \sum_{i=1}^N \frac{\|\mathbf{a}_i\|^2}{2} \left(|x_i| - \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|} \right)^2 \mathbb{1}_{\{|x_i| \leq \frac{\sqrt{2\lambda}}{\|\mathbf{a}_i\|}\}}. \quad (11)$$

In (11), \mathbf{a}_i denotes the i -th column of \mathbf{A} and x_i the i -th element of \mathbf{x} . The resulting ℓ_2 -CEL0 functional (10) preserves global minimizers of the initial ℓ_2 - ℓ_0 criteria (7) while some local minimizers are removed. Hence, solving (10) instead of (7) allows a better minimization of ℓ_2 - ℓ_0 by avoiding some of its local minimizers. Again, (10) can be minimized through FBS¹ where the proximity operator used in (4) is given, in the case of the CEL0 penalty, by a new thresholding rule defined by

$$\text{prox}_{\gamma \Phi_{\text{CEL0}}(\cdot)}(\mathbf{u}) = [K_{\text{CEL0}}(u_i)]_{i \in \{1, \dots, N\}}, \quad (12)$$

with,

$$K_{\text{CEL0}}(u_i) = \frac{u_i}{|u_i|} \min \left(|u_i|, \frac{\left(|u_i| - \sqrt{2\lambda\gamma} \|\mathbf{a}_i\| \right)_+}{1 - \|\mathbf{a}_i\|^2 \gamma} \right), \quad (13)$$

where u_i denotes the i -th element of \mathbf{u} . As can be seen, the threshold in CEL0 penalization is different for each element in \mathbf{u} as it depends on $\|\mathbf{a}_i\|$ and is continuous.

In the following we will compare for channel and DOA estimation, the performances of the minimization problems ℓ_2 - ℓ_0 (7), ℓ_2 -CEL0 (10) and ℓ_2 - ℓ_1 , by using the same FBS algorithm leading to use hard threshold (8), CEL0 threshold (13) and soft threshold (5).

¹with the nonconvex case convergence result [13].

IV. ℓ_0 -OPTIMIZATION IN CHANNEL ESTIMATION

In this section, we address the problem of sparse channel estimation by the iterative methods presented previously. In wireless communications, due to multipath effect, the signal at the receiver is a composition of scaled and delayed versions of the transmitted signal. The channel can be modeled as a finite set of K scaled delays applied to the transmitted signal. Considering that the K delays can be distributed on a grid of $N > K$ positions spaced by T_G seconds and that the signal can be represented in discrete time domain with a sample rate of $1/T_G$, the problem can be written in its matricial form as $\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{n}$, where $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_M]^T$ are the observations, $\mathbf{h} = [h_1 \ h_2 \ \dots \ h_N]^T$ are the unknown channel taps to be estimated, \mathbf{n} is the noise vector and

$$\mathbf{S} = \begin{bmatrix} s_N & s_{N-1} & \dots & s_1 \\ s_{N+1} & s_N & \dots & s_2 \\ \vdots & \vdots & \ddots & \vdots \\ s_{N+M+1} & s_{N+M} & \dots & s_M \end{bmatrix} \quad (14)$$

is part of the transmitted information. When the signal is pilot-aided, the symbols s_k are known at receiver enabling the possibility of channel estimation based on the information received. However, if the sample frequency at the receiver is lower than $1/T_G$ or if signal's bandwidth is lower than $1/(2T_G)$, channel estimation through traditional methods are unreliable due to the correlation between the columns of \mathbf{S} . Additionally, if the channel taps are very close, as \mathbf{S} has high mutual coherence, ℓ_1 relaxation delivers bad precision and resolution and ℓ_0 -optimization should be used instead.

To exemplify ℓ_0 -optimization in this case, a set of simulations was performed for penalized problem (7). A 3GPP-LTE ETU channel model [14] was considered. The channel taps were distributed over a 600 equispaced tap grid, with space between taps of 10 ns. A pilot-aided signal was generated, passed through the channel and corrupted by gaussian random noise. The pilot size was chosen to be 512 symbols and the spectral bandwidth of the signal was considered to be 50 MHz. Finally, the sample rate at receiver was configured to 33.3 MHz. Note that this configuration violates the Shannon-Nyquist criterium. Additionally, due to the signal's spectral bandwidth, a symbol spans over two or more grid taps, producing a high-correlated mixture matrix.

Figure 1 shows the estimation error due to ℓ_1 relaxed optimization performed with FBS and when estimation is performed using ℓ_0 -optimization through IHT and FBS with CEL0 penalization. All the optimization algorithms used are initialized with $\hat{\mathbf{h}}^{(0)} = \mathbf{0}$ and the parameter λ is, for each case, the one providing the best performance (lowest MSE) over a range of tested λ values. Note that the same value of λ is used for all SNR levels. As reference, the Oracle-CRB [16], i.e. the minimal error that can be obtained knowing the support in function of noise, is shown. Clearly, ℓ_0 -optimization performs better than ℓ_1 -optimization in this case. Moreover, CEL0 penalization improves the estimation in noisy environments when compared with traditional IHT approach.

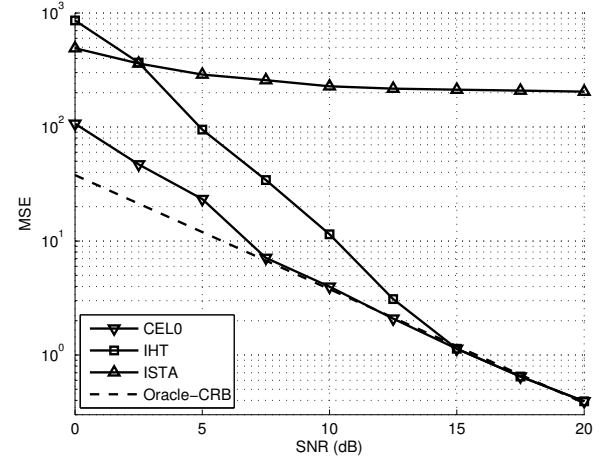


Figure 1. MSE vs. SNR for channel estimation using ℓ_1 minimization (FBS with soft threshold) and ℓ_0 minimization (IHT or CEL0 with FBS).

V. ℓ_0 -OPTIMIZATION IN DOA ESTIMATION

In this section we deal with Uniform Linear Array (ULA) compound by isotropic sensors. We assume that K i.i.d. narrowband signals impinge an M -element antenna array, whose outputs are sampled, generating L snapshots. For $L > 1$ the problem can be written in matrix form as $\mathbf{Y} = \mathbf{A}\mathbf{S} + \mathbf{N}$, where $\mathbf{Y} \in \mathbb{C}^{M \times L}$ is the matrix of measurements, $\mathbf{S} \in \mathbb{C}^{K \times L}$ the signal matrix, $\mathbf{N} \in \mathbb{C}^{M \times L}$ the noise matrix and $\mathbf{A} \in \mathbb{C}^{M \times K}$ is the mixture matrix defined as $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_K]^T$, where

$$\mathbf{a}_i = [1 \ e^{j\pi \sin \phi_i} \ e^{j2\pi \sin \phi_i} \ \dots \ e^{j(K-1)\pi \sin \phi_i}]^T \quad (15)$$

is the i -th impinging signal steering vector with incident angle $\phi_i \in [-90^\circ, +90^\circ]$. In (15), the elements are considered spaced by half of the wavelength.

Here, the mixture matrix is not known, and in the sense of sparse optimization, an overcomplete representation of \mathbf{A} is considered. In this way, the continuous range of possible incident angles is sliced in N taps of $\Delta\phi^\circ$ to obtain a new mixture matrix $\mathbf{A}_G \in \mathbb{C}^{M \times N}$ which can then be used to estimate a highly-sparse matrix $\mathbf{S}_G \in \mathbb{C}^{N \times L}$ that has only K non-zero lines. When $L = 1$ we can directly apply ℓ_0 -optimization through IHT or FBS with CEL0 penalization as done for channel estimation. However, when multiple snapshots are considered, we aim to find a *row-sparse* estimate $\hat{\mathbf{S}}_G$. Hence, previously presented methods have to be extended to the Multiple Measurement Vector (MMV) case. This can be easily achieved by modifying the thresholding rules (8) and (13) considering u_i as a vector (here a row of \mathbf{S}_G) and replacing the absolute value $|\cdot|$ by the ℓ_2 -norm (as in [17]).

In order to illustrate ℓ_0 -optimization applied to DOA estimation, we consider $K = 2$ equipower signals impinging an $M = 10$ elements ULA under directions of arrival $\phi_1 = 0^\circ$ and $\phi_2 = 5^\circ$. The observations \mathbf{Y} were corrupted by zero mean gaussian noise. The estimation grid step was chosen to be 0.15° and the interval of possible incident angles to be estimated was considered to be $[-45^\circ, +45^\circ]$. Two situations regarding the number of snapshots were considered: $L = 1$ and

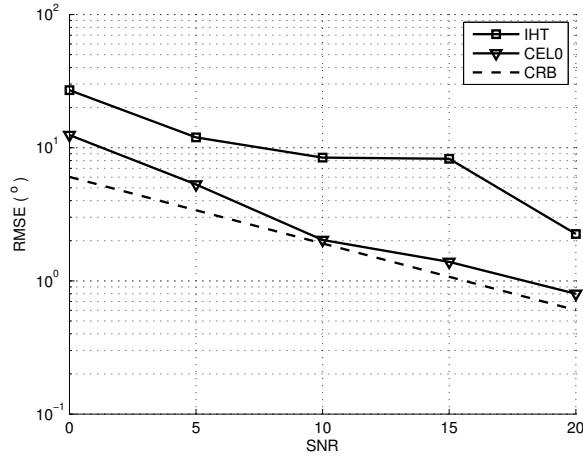


Figure 2. RMSE vs. SNR for DOA estimation using ℓ_0 -optimization for $M = 10$ and $L = 1$ snapshots.

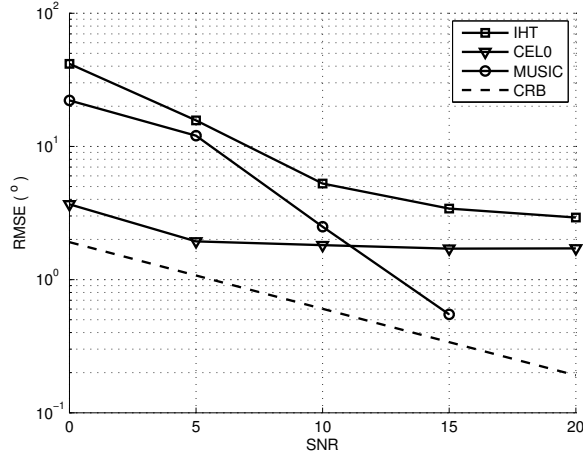


Figure 3. RMSE vs. SNR for DOA estimation using ℓ_0 -optimization and MUSIC [15] for $M = 10$ and $L = 10$ snapshots.

$L = 10$. In both cases 100 Monte-Carlo runs were performed and the results are compared in terms of the root-mean-square error (RMSE) in degrees between the supports of \mathbf{S}_G and $\hat{\mathbf{S}}_G$. Algorithms initialization and hyperparameter (λ) selection are the same as for the channel estimation problem. It is important to note that in both cases the mixture matrix \mathbf{A}_G is extremely correlated, prohibiting the use of ℓ_1 relaxation.

Figure 2 depicts the RMSE achieved with IHT and CEL0 when $L = 1$. ℓ_1 relaxation fails in such a severe scenario. As can be seen, CEL0 penalization improves the performance of IHT no matter how low is the SNR. It is interesting to note too that results for CEL0 are very close to the Cramer-Rao lower bound (CRB). Figure 3 shows the simulation results when $L = 10$, i.e. considering the MMV case. Results using MUSIC algorithm [15] are also provided as a reference. At low SNR, ℓ_0 -optimization algorithms present performance similar to MUSIC and even outperform it with CEL0 penalization. This is an interesting result in regards to the fact that MUSIC knows the number of sources (as CRB does), which is not the case for IHT and CEL0. For high SNR, the MUSIC algorithm

leads to better results.

VI. CONCLUSION

In this paper we have shown how nonconvex regularization terms with dedicated penalized algorithms recently introduced in signal and image literature can be successful for DOA and wireless channel estimation. It is shown that nonconvex penalties outperform classical ℓ_1 relaxation in these two classical problems in array processing, at the same computational cost. The continuous CEL0 penalization improves results when using FBS algorithm w.r.t. to IHT algorithm. Future work will consider constrained formulation of ℓ_0 optimization in the case of known sparsity (number of taps for channel estimation and number of sources for DOA estimation).

ACKNOWLEDGMENT

The authors would like to thank the NoE Newcom# under contract FP7-ICT-318306 and the CAPES-COFECUB Ph 772-13 “CompEst” project.

REFERENCES

- [1] E.J. Candès and M.B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, 25 (2008), pp. 21-30.
- [2] B.K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Comput.*, 24 (1995), pp. 227-234.
- [3] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, 20 (1998), pp. 33-61.
- [4] I. Daubechies, M. Defrise, and C. de Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Comm. Pure Appl. Math.*, 57 (2004), pp. 1413-1457.
- [5] S.G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Process.*, 41 (1993), pp. 3397-3415.
- [6] J. Tropp and A.C. Gilbert, “Signal recovery from partial information via orthogonal matching pursuit,” *IEEE Trans. Information Theory*, 53 (2007), pp. 4655-4666.
- [7] C. Soussen, J. Idier, D. Brie and J. Duan, “From Bernoulli-Gaussian deconvolution to sparse signal restoration,” *IEEE Trans. Signal Process.*, 59 (2011), pp. 4572-4584.
- [8] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Information Theory*, 51 (2005), pp. 4203-4215.
- [9] T. Blumensath and M. E. Davies, “Iterative Thresholding for Sparse Approximations,” *J. Fourier Anal. Appl.*, 14 (2008), pp. 629-654.
- [10] E. Soubies, L. Blanc-Féraud, and G. Aubert, “A Continuous Exact ℓ_0 penalty (CEL0) for least squares regularized problem,” *SIAM J. Imaging Sci.*, 8 (2015), pp. 1574-1606.
- [11] P.L. Combettes and V.R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Model. Simul.* 4 (2005), pp. 1168-1200.
- [12] J.J. Moreau, “Proximité et dualité dans un espace hilbertien,” *Bull. Soc. Math.*, France 93, 273299, 1965.
- [13] H. Attouch, J. Bolte and B. F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods,” *Math. Program.*, 137 (2013), pp. 91-129.
- [14] 3rd Generation Partnership Project, Tech. Specification Group Radio Access Network, Evolved Universal Terrestrial Radio Access (E-UTRA), “Physical Channels and Modulation,” Release 9, V9.1.0, Mar. 2010.
- [15] R. O. Schmidt, “Multiple Emitter Location and Signal Parameter Estimation,” *IEEE Trans. Antennas Propagation*, AP-34 (1986), pp. 276-280.
- [16] Z. Ben-Haim & Y. C. Eldar, “The Cramer-Rao Bound for Estimating a Sparse Parameter Vector,” *IEEE Trans. Signal Process.*, 58 (2010), pp. 3384-3389.
- [17] C. Stoeckle, J. Munir, A. Mezghani & J.A. Nossek, “DoA Estimation Performance and Computational Complexity of Subspace- and Compressed Sensing-based Methods,” *International ITG Workshop on Smart Antennas* (2015), pp. 1-6.